

THE MEDIAN LARGEST PRIME FACTOR AND THE MEAN OF $\omega(n)$

ERIC NASLUND

ABSTRACT. In [4], Diaconis used contour integration to give an asymptotic expansion of $\sum_{n \leq x} \omega(n)$ up to an error of $O\left(\frac{x}{(\log x)^k}\right)$ for any integer k , where $\omega(n) = \sum_{p|n} 1$ denotes the number of distinct prime factors function. In this article, we use the hyperbola method to provide a more precise formulation with an error term of the form $O\left(xe^{-c\sqrt{\log x}}\right)$, and deduce the expansion in [4] as a consequence. We also show that Diaconis' expansion follows from the earlier work of De Bruijn [2] on integers without large prime factors. Using this expansion, we then prove that if $M(x)$ denotes the median largest prime factor of the integers in the interval $[1, x]$, we have

$$M(x) = e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}} \left(1 + O\left(\frac{1}{\log x}\right)\right),$$

where γ is the Euler Mascheroni constant. Our method is then applied in an identical way to obtain an asymptotic expansion for the average largest prime factor, $\sum_{n \leq x} P(n)$, reproving results of [1, 3], with the advantage that the constants in the expansion are given explicitly, and an integral function is given with a more precise error term.

1. INTRODUCTION

The number of distinct prime divisors function, $\omega(n)$, has been the subject of study for centuries. In 1917 Hardy and Ramanujan proved that $\omega(n)$ has normal order $\log \log n$, that is $\omega(n) \sim \log \log n$ for almost all integers n . Turán reproved this result in an elementary way by showing that both the mean and variance of $\omega(n)$ on $[1, x]$ are asymptotic to $\log \log x$. [7] In 1976 Diaconis used contour integration to give an asymptotic expansion for the mean and variance of $\omega(n)$ [4], and in particular for the mean he proved that for any integer k ,

$$(1.1) \quad \sum_{n \leq x} \omega(n) = -x \log \log x + B_1 x + c_0 \frac{x}{\log x} + c_1 \frac{x}{\log^2 x} + \cdots + c_{k-1} \frac{(k-1)!x}{\log^k x} + O\left(\frac{x}{\log^{k+1} x}\right),$$

for some computable constants c_k . In section 2, we use the hyperbola method to prove that the following theorem:

Theorem 1. *Letting $li_f(x) = \int_2^x \frac{\{t\}}{\log t} dt$, we have that*

$$\sum_{n \leq x} \omega(n) = x \log \log x + B_1 x - li_f(x) + O\left(xe^{-c\sqrt{\log x}}\right),$$

and for any integer k

$$(1.2) \quad li_f(x) = c_0 \frac{x}{\log x} + c_1 \frac{x}{\log^2 x} + \cdots + c_{k-1} \frac{(k-1)!x}{\log^k x} + O\left(\frac{x}{\log^{k+1} x}\right)$$

where $c_n = 1 - \sum_{k=0}^n \frac{1}{k!} \gamma_k$, γ_k denotes the k^{th} Stieltjes constant, and $c_n \rightarrow \frac{1}{2}$ as $n \rightarrow \infty$.

From this theorem equation 1.1 follows as a corollary, and in section 3 we use the results of [2] and [6] to provide an additional proof of this result. With this expansion, we prove the following regarding the median prime factor:

Theorem 2. *If $M(x)$ is the median prime factor of the interval $[1, x]$, then*

$$M(x) = e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}} \left(1 + O\left(\frac{1}{\log x}\right) \right).$$

In section 3, using Theorem 1, a more quantitative version of this theorem is given in terms of the function $\text{li}_f(x)$ with an error term of the form $e^{-c\sqrt{\log x}}$.

In our final section, we examine $\sum_{n \leq x} P(n)$, the average of the largest prime factor of n . This was first looked at by Erdős and Alladi in [1], and they proved that $\sum_{n \leq x} P(n) \sim \frac{\pi^2}{12} \frac{x^2}{\log x}$. Later, in [3], De Koninck and Ivić showed that $\sum_{n \leq x} P(n)$ has an asymptotic expansion with computable constants. Using the method of proof of Theorem 1, we are able to prove the following:

Theorem 3. *Letting $\text{li}_g(x) = \int_2^x \frac{t}{x} \frac{[\frac{x}{t}]}{\log t} dt$, we have that*

$$\sum_{n \leq x} P(n) = x \text{li}_g(x) + O\left(x^2 e^{-c\sqrt{\log x}}\right).$$

and for any integer k

$$(1.3) \quad \text{li}_g(x) = d_0 \frac{x}{\log x} + d_1 \frac{1!x}{(\log x)^2} + \cdots + d_{k-1} \frac{(k-1)!x}{(\log x)^k} + O\left(\frac{x}{(\log x)^{k+1}}\right)$$

where $d_n = \frac{1}{2^{n+1}} \sum_{j=0}^n \frac{2^j (-1)^j \zeta^{(j)}(2)}{j!}$, with $\zeta^{(j)}(2)$ denoting the j^{th} derivative of $\zeta(s)$ evaluated at 2, and $d_n \rightarrow 1$ as $n \rightarrow \infty$.

Notice in particular that $d_0 = \frac{\pi^2}{12}$, so we deduce that

$$\sum_{n \leq x} P(n) = \frac{\pi^2}{12} \frac{x^2}{\log x} + O\left(\frac{x^2}{(\log x)^2}\right).$$

Although this expansion has previously been presented in [3], and again in [5], there are some advantages to the result above. We give an explicit form for the constants in terms of the zeta function, and our proof is basic requiring only one application of the hyperbola method. Also, we give the explicit integral form which has an error like that of the prime number theorem mimicking how $\pi(x)$ is approximated by the function $\text{li}(x)$.

It is worth mentioning why the notation $\text{li}_f(x), \text{li}_g(x)$ is used for the above integral functions. This is because not only do they mimic $\text{li}(x) = \int_2^x \frac{1}{\log t} dt$ in terms of their definition and the error terms which appear, but also because they have a nearly identical asymptotic expansion. Recall that for $\text{li}(x) = \int_2^x \frac{1}{\log t} dt$, by integration by parts we have that

$$(1.4) \quad \text{li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2!x}{\log^3 x} + \cdots + \frac{(k-1)!x}{\log^k x} + O\left(\frac{x}{\log^{k+1} x}\right).$$

This is very similar to both equations 1.2 and 1.3, and the similarities are made clearer by the fact that as $n \rightarrow \infty$,

$$c_n \rightarrow \frac{1}{2}, \text{ and } d_n \rightarrow 1.$$

2. THE MEAN VALUE OF $\omega(n)$ AND $\Omega(n)$.

Our goal is to find a more precise asymptotic formula for $\sum_{n \leq x} \omega(n)$ and $\sum_{n \leq x} \Omega(n)$. We start by looking at the sum of $\omega(n)$. Since $\omega(n) = \sum_{p|n} 1$, by rearranging, we have that

$$\sum_{n \leq x} \omega(n) = \sum_{n \leq x} \sum_{p|n} 1 = \sum_{p \leq x} \left[\frac{x}{p} \right] = x \sum_{p \leq x} \frac{1}{p} - \sum_{p \leq x} \left\{ \frac{x}{p} \right\}.$$

The prime number theorem tells us that

$$(2.1) \quad \sum_{p \leq x} \frac{1}{p} = \log \log x + B_1 + O\left(e^{-c\sqrt{\log x}}\right)$$

where $B_1 = \gamma - \sum_p \sum_{k \geq 2} \frac{1}{kp^k}$ is Merten's constant, so we conclude that

$$\sum_{n \leq x} \omega(n) = x \log \log x + B_1 x - \sum_{p \leq x} \left\{ \frac{x}{p} \right\} + O\left(xe^{-c\sqrt{\log x}}\right).$$

From here we need only evaluate the sum of fractional parts over the primes. Since the primes have density $\frac{1}{\log x}$ we expect that $\sum_{p \leq x} \left\{ \frac{x}{p} \right\} \approx \int_2^x \left\{ \frac{x}{t} \right\} \frac{dt}{\log t}$, and this is indeed the case. The error term is managed by using the hyperbola method along with the prime number theorem, and we have the following proposition:

Proposition 4. *Letting $li_f(x) := \int_2^x \frac{\{x/t\}}{\log t} dt$, we have that*

$$\sum_{p \leq x} \left\{ \frac{x}{p} \right\} = li_f(x) + O\left(xe^{-c\sqrt{\log x}}\right).$$

Proof. Let $1 < B \leq x$ be some integer. Splitting into intervals and rearranging the order, we have that

$$\begin{aligned} \sum_{\frac{x}{B} < p \leq x} \left[\frac{x}{p} \right] &= \sum_{n \leq B-1} n \left(\sum_{\frac{x}{n+1} < p \leq \frac{x}{n}} 1 \right) = \pi(x) + \pi\left(\frac{x}{2}\right) + \cdots + \pi\left(\frac{x}{B-1}\right) - (B-1)\pi\left(\frac{x}{B}\right) \\ (2.2) \quad &= \sum_{n \leq B-1} \left(\pi\left(\frac{x}{n}\right) - \pi\left(\frac{x}{B}\right) \right). \end{aligned}$$

By the prime number theorem this is

$$\begin{aligned} &= \sum_{n \leq B-1} \int_{\frac{x}{B}}^{\frac{x}{n}} \frac{1}{\log t} dt + O\left(\sum_{n \leq B-1} \left(\frac{x}{n} e^{-c\sqrt{\log \frac{x}{n}}}\right)\right) \\ &= \int_{\frac{x}{B}}^x \frac{[x/t]}{\log t} dt + O\left(xe^{-c\sqrt{\log \frac{x}{B}}} \log B\right). \end{aligned}$$

Splitting up the floor function, the main term is

$$\int_{\frac{x}{B}}^x \frac{[x/t]}{\log t} dt = x \left(\log \log x - \log \log \left(\frac{x}{B}\right) \right) - \int_{\frac{x}{B}}^x \frac{\{x/t\}}{\log t} dt,$$

and hence since $\sum_{\frac{x}{B} < p \leq x} \frac{1}{p} = \log \log x - \log \log \left(\frac{x}{B}\right) + O\left(e^{-c\sqrt{\log x}}\right)$ by 2.1, we have that

$$\sum_{\frac{x}{B} < p \leq x} \left\{ \frac{x}{p} \right\} = \int_{\frac{x}{B}}^x \frac{\{x/t\}}{\log t} dt + O\left(xe^{-c\sqrt{\log \frac{x}{B}}} \log B\right).$$

The result then follows by choosing $B = \sqrt{x}$ and noting that we can extend the sum and integral to start at 2 as $\int_2^{\sqrt{x}} \frac{\{x/t\}}{\log t} dt = O\left(\frac{\sqrt{x}}{\log x}\right)$ and $\sum_{p \leq \sqrt{x}} \left\{\frac{x}{p}\right\} = O\left(\frac{\sqrt{x}}{\log x}\right)$. \square

Corollary 5. *We have that*

$$\sum_{n \leq x} \omega(n) = x \log \log x + B_1 x - li_f(x) + O\left(xe^{-c\sqrt{\log x}}\right)$$

and

$$\sum_{n \leq x} \Omega(n) = x \log \log x + B_2 x - li_f(x) + O\left(xe^{-c\sqrt{\log x}}\right)$$

where

$$B_2 = B_1 + \sum_p \frac{1}{p(p-1)}.$$

Proof. The first identity is immediate from what has been done thus far. As $\sum_{n \leq x} \Omega(n) = \sum_{p \leq x} \sum_{k \geq 1} \left\lfloor \frac{x}{p^k} \right\rfloor$ we note that difference is given explicitly by

$$\sum_{n \leq x} \Omega(n) - \omega(n) = \sum_{p \leq x} \sum_{k \geq 2} \left\lfloor \frac{x}{p^k} \right\rfloor = \sum_{2 \leq k \leq \log_2(x)} \sum_{p \leq \sqrt{x}} \left\lfloor \frac{x}{p^k} \right\rfloor = x \sum_{2 \leq k \leq \log_2(x)} \sum_{p \leq \sqrt{x}} \frac{1}{p^k} + O(\sqrt{x} \log x).$$

Extending the sums to ∞ and rearranging we have $\sum_{n \leq x} \Omega(n) - \omega(n) = x \sum_p \frac{1}{p(p-1)} + O(\sqrt{x} \log x)$ which implies the result. \square

2.1. The Function $li_f(x)$. In this subsection we provide an asymptotic expansion for $li_f(x)$. We will make use of the laurent expansion of $\zeta(s)$ which is given by

$$(2.3) \quad \zeta(s) = \frac{1}{s-1} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma_n (s-1)^n.$$

where the γ_n are the Stieltjes Constants, and γ_0 is the Euler-Mascheroni constant. The expansion for $li_f(x)$ can then be expressed in terms of these constants.

Proposition 6. *For any integer $k > 0$ we have*

$$(2.4) \quad li_f(x) = c_0 \frac{x}{\log x} + c_1 \frac{x}{\log^2 x} + c_2 \frac{2!x}{\log^3 x} + \cdots + c_{k-1} \frac{(k-1)!x}{\log^k x} + O\left(\frac{x}{\log^{k+1} x}\right)$$

where

$$c_n = 1 - \sum_{k=0}^n \frac{1}{k!} \gamma_k.$$

By using the fact that $\zeta(0) = -\frac{1}{2}$ along with Abel's Limit Theorem we find that $\sum_{k=0}^{\infty} \frac{1}{k!} \gamma_k = \frac{1}{2}$ since this series is absolutely convergent. This tells us that as $n \rightarrow \infty$, $c_n \rightarrow \frac{1}{2}$ which is roughly what we might expect since $\int_2^x \{t\} dt \approx \frac{x}{2}$. From Proposition 6 along with Corollary 5 we are able to deduce Theorem 1. To prove this proposition recall that for all $s \neq 1$, $\text{Re}(s) > 0$, we have

$$(2.5) \quad \zeta(s) = \frac{1}{s-1} + 1 - s \int_1^{\infty} \{x\} x^{-s-1} dx,$$

which allows us to prove the following lemma:

Lemma 7. For any integer $n \geq 0$ we have

$$\int_1^\infty \frac{\{u\}}{u^2} (\log u)^n du = n! \left(1 - \sum_{k=0}^n \frac{1}{k!} \gamma_n \right) = n! c_n$$

Proof. Let $b_n = \int_1^\infty \frac{\{u\}}{u^2} (\log u)^n du$. Then to find each b_n , consider the generating series

$$\sum_{n=0}^\infty \frac{b_n}{n!} x^n = \int_1^\infty \frac{\{u\}}{u^2} e^{x \log u} du = \int_1^\infty \{u\} u^{x-2} du.$$

By 2.5, this equals $\frac{1-\zeta(1-x)-\frac{1}{x}}{1-x}$. Using 2.3 we then have

$$\sum_{n=0}^\infty \frac{b_n}{n!} x^n = \left(\sum_{m \geq 0} x^m \right) \left(1 - \sum_{n \geq 0} \frac{\gamma_n}{n!} x^n \right)$$

and upon comparing coefficients, the result follows. \square

The proposition now follows from some basic substitutions:

Proof. In the equation

$$(2.6) \quad \text{li}_f(x) = \int_{\sqrt{x}}^x \frac{\{x/t\}}{\log t} dt + O\left(\frac{\sqrt{x}}{\log x}\right),$$

let $u = x/t$. This allows us to rewrite the integral as

$$\int_{\sqrt{x}}^x \frac{\{x/t\}}{\log t} dt = x \int_1^{\sqrt{x}} \frac{\{u\}}{u^2 \log\left(\frac{x}{u}\right)} du = \frac{x}{\log x} \int_1^{\sqrt{x}} \frac{\{u\}}{u^2} \left(1 - \frac{\log u}{\log x}\right)^{-1} du.$$

Let $k > 0$ be some integer, and expand the geometric series

$$\left(1 - \frac{\log u}{\log x}\right)^{-1} = 1 + \frac{\log u}{\log x} + \cdots + \left(\frac{\log u}{\log x}\right)^{k-1} + \left(\frac{\log u}{\log x}\right)^k \left(1 - \frac{\log u}{\log x}\right)^{-1}$$

so that

$$\int_{\sqrt{x}}^x \frac{\{x/t\}}{\log t} dt = \frac{x}{\log x} \sum_{n=0}^{k-1} \frac{\int_1^{\sqrt{x}} \frac{\{u\}}{u^2} (\log u)^n du}{(\log x)^n} + \frac{x}{(\log x)^{k+1}} \int_1^{\sqrt{x}} \frac{\{u\}}{u^2} (\log u)^k \left(1 - \frac{\log u}{\log x}\right)^{-1} du.$$

Since $\log u \leq \frac{1}{2} \log x$, it follows that

$$\int_1^{\sqrt{x}} \frac{\{u\}}{u^2} (\log u)^k \left(1 - \frac{\log u}{\log x}\right)^{-1} du \leq 2 \int_1^\infty \frac{\{u\}}{u^2} (\log u)^k du < \infty$$

so the last term contributes an error of the form $O_k\left(\frac{x}{\log^{k+1} x}\right)$. Since we may bound the integral

$$\int_{\sqrt{x}}^\infty \frac{\{u\}}{u^2} (\log u)^n du \leq \int_{\sqrt{x}}^\infty \frac{(\log u)^n}{u^2} du = O\left(\frac{(\log x)^n}{\sqrt{x}}\right)$$

it follows that

$$(2.7) \quad \int_{\sqrt{x}}^x \frac{\{x/t\}}{\log t} dt = \frac{x}{\log x} \sum_{n=0}^{k-1} \frac{\int_1^{\sqrt{x}} \frac{\{u\}}{u^2} (\log u)^n du}{(\log x)^n} + O_k\left(\frac{x}{\log^{k+1} x}\right).$$

The stated result then follows by combining equation 2.6 and 2.7 along with Lemma 7. \square

3. THE MEDIAN LARGEST PRIME FACTOR

Let $\psi(x, y)$ denote the number of integers n with $1 \leq n \leq x$, all of whose prime factors are $\leq y$. Then the median prime factor $M(x)$ for the integers in the interval $[1, x]$ satisfies

$$\psi(x, M(x)) = \frac{1}{2}x.$$

Our main result is the following:

Theorem 8. *We have that*

$$M(x) = x^{\frac{1}{\sqrt{e}} \exp\left(-\frac{\text{li}_f(x)}{x}\right)} + O\left(x^{\frac{1}{\sqrt{e}}} e^{-c\sqrt{\log x}}\right).$$

Before proving Theorem 8, we show that Theorem 2 follows from the above along with Proposition 6, the asymptotic expansion for $\text{li}_f(x)$.

Proof. Since

$$\frac{\text{li}_f(x)}{x} = \frac{1-\gamma}{\log x} + O\left(\frac{1}{\log^2 x}\right),$$

by using the series expansion for e^x we see that

$$\exp\left(-\frac{\text{li}_f(x)}{x}\right) = 1 - \frac{1-\gamma}{\log x} + O\left(\frac{1}{\log^2 x}\right).$$

From here, the series expansion for e^x applied once again tells us that

$$(3.1) \quad x^{\frac{1}{\sqrt{e}} \exp\left(-\frac{\text{li}_f(x)}{x}\right)} = e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}} + O\left(\frac{x^{\frac{1}{\sqrt{e}}}}{\log x}\right),$$

proving the desired result. □

Let $\psi(x, y)$ denote the number of integers n with $1 \leq n \leq x$, all of whose prime factors are $\leq y$. If $x^{1/2} \leq y \leq x$, then $n \leq x$ implies that at most one prime $p > y$ can divide n , so we have that

$$\psi(x, y) = [x] - \sum_{y < p \leq x} \sum_{\substack{n \leq x \\ p|n}} 1 = [x] - \sum_{y < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor.$$

Since

$$\sum_{y < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor = x \sum_{y < p \leq x} \frac{1}{p} - \sum_{y < p \leq x} \left\{ \frac{x}{p} \right\},$$

we see that

$$(3.2) \quad \psi(x, y) = x - \sum_{n \leq x} \omega(n) + \sum_{p \leq y} \frac{1}{p} + O\left(\pi(y) + x e^{-c\sqrt{\log x}}\right)$$

so that by Theorem 4 we have

$$(3.3) \quad \psi(x, y) = x \left(1 - \log\left(\frac{\log x}{\log y}\right)\right) + \text{li}_f(x) + O\left(\pi(y) + x e^{-c\sqrt{\log x}}\right).$$

In the next subsection, we will use equation 3.2 and a result of De Bruijn to provide an alternate derivation for the asymptotic expansion of $\sum_{n \leq x} \omega(n)$ given by Diaconis. With equation 3.3 we are now ready to prove Theorem 8.

Proof. Let $M(x)$ be the median prime factor for the interval $[1, x]$ so that

$$\psi(x, M(x)) = \frac{1}{2}x.$$

Then by equation 3.3, we must have that

$$\frac{1}{2}x = x \left(1 - \log \left(\frac{\log x}{\log M(x)} \right) \right) + \text{li}_f(x) + O \left(x e^{-c\sqrt{\log x}} \right),$$

and hence

$$-\log \left(\frac{\log x}{\sqrt{e} \log M(x)} \right) = -\frac{\text{li}_f(x)}{x} + O \left(e^{-c\sqrt{\log x}} \right).$$

Taking the exponential of both sides, and noting that $\exp \left(O \left(e^{-c\sqrt{\log x}} \right) \right) = 1 + O \left(e^{-c\sqrt{\log x}} \right)$, we find that

$$\frac{\sqrt{e} \log M(x)}{\log x} = \exp \left(-\frac{\text{li}_f(x)}{x} \right) + O \left(e^{-c\sqrt{\log x}} \right).$$

Multiplying through by $\frac{\log x}{\sqrt{e}}$ and taking the exponential once again, we find that

$$M(x) = x^{\frac{1}{\sqrt{e}} \exp \left(-\frac{\text{li}_f(x)}{x} \right)} + O \left(x^{\frac{1}{\sqrt{e}}} e^{-c\sqrt{\log x}} \right),$$

proving the result. \square

3.1. Integers without large prime factors: De Bruijn's result. In [2], De Bruijn showed that

$$\psi(x, y) = \Lambda(x, y) + O \left(x e^{-c\sqrt{\log y}} \right)$$

where

$$\Lambda(x, y) = x \int_0^\infty \rho \left(\frac{\log x - \log t}{\log y} \right) d \frac{[t]}{t}$$

and $\rho(u)$ denotes the Dickmann De Bruijn rho function. The error term in the above was later improved by Saias [6]. In his paper, De Bruijn also gave an asymptotic expansion for $\Lambda(x, y)$, in terms of computable constants, however we will use the version appearing in Saias' paper as it is easier to work with. Specifically Saias shows that if $u = \frac{\log x}{\log y}$ so that $x^{\frac{1}{u}} = y$, and $u \notin \mathbb{N}$, $u \leq (\log y)^{\frac{3}{5}-\epsilon}$, then we have

$$(3.4) \quad \Lambda(x, y) = x \sum_{k=0}^n a_k \frac{\rho^{(k)}(u)}{(\log y)^k} + O_{n,\epsilon} \left(x \frac{\rho^{(n+1)}(u)}{(\log y)^{n+1}} \right)$$

where

$$a_0 = 1, \quad a_k = \frac{(-1)^k}{(k-1)!} \int_1^\infty \frac{\{t\}}{t^2} (\log t)^{k-1} dt.$$

In what follows, we will have no use for the range on u , as what we are interested in is when $1 < u < 2$. On this range, we have that $\rho(u) = 1 - \log u$ so that $\rho'(u) = -\frac{1}{u}$ and

$$\rho^{(k)}(u) = \frac{(k-1)!(-1)^k}{u^k}.$$

We note that, by 7, for $k \geq 1$

$$a_k = (-1)^k c_{k-1} = (-1)^k \left(1 - \sum_{j=0}^{k-1} \frac{1}{j!} \gamma_j \right).$$

Since $\frac{1}{(\log y)^k} = \frac{u^k}{(\log x)^k}$, equation 3.4 becomes

$$(3.5) \quad \Lambda(x, y) = x\rho(u) + \sum_{k=1}^n \frac{(k-1)!c_{k-1}}{(\log x)^k} + O_n\left(\frac{x}{(\log x)^{n+1}}\right),$$

and hence using the fact that for $1 < u < 2$ we have $\rho(u) = 1 - \log u = 1 - \log \log x - \log \log y$, and by 3.2 conclude that

$$\sum_{n \leq x} \omega(n) = \log \log x + B_1 - \sum_{k=0}^{n-1} \frac{(k-1)!c_k}{(\log x)^{k+1}} + O_n\left(\frac{x}{(\log x)^{n+1}}\right).$$

This reproves 1.1, Diaconis' expansion.

Remark 9. Rather than providing the elementary derivation of Corollary 5, and using this to prove Theorem 2, we could have proceeded directly from the results of Saias and De Bruijn using equation 3.5 to prove Theorem 2. One of the goals the author had in writing this was to show how all of these things are interrelated and equivalent.

4. THE AVERAGE LARGEST PRIME FACTOR

In this section, following the proof of Theorem 1 very closely, we will deduce Theorem 3. First, note that each integer n may have at most one prime $p > \sqrt{x}$ such that $p|n$, and for each prime p , there are $\left[\frac{x}{p}\right]$ integers divisible by p which are less than x . Combining these two facts, we see that

$$\sum_{\sqrt{x} \leq p \leq x} p \left[\frac{x}{p}\right] \leq \sum_{n \leq x} P(n) \leq \sum_{p \leq x} p \left[\frac{x}{p}\right],$$

and since $\sum_{p \leq \sqrt{x}} p \left[\frac{x}{p}\right] \leq x \sum_{p \leq \sqrt{x}} 1 = O\left(\frac{x^{\frac{3}{2}}}{\log x}\right)$, it follows that

$$\sum_{n \leq x} P(n) = \sum_{\sqrt{x} < p \leq x} p \left[\frac{x}{p}\right] + O\left(\frac{x^{\frac{3}{2}}}{\log x}\right).??$$

Define $\mathcal{K}(x) = \sum_{p \leq x} p$, so that by the prime number theorem along with partial summation, we have the estimate

$$(4.1) \quad \mathcal{K}(x) = \int_2^x \frac{t}{\log t} dt + O\left(x^2 e^{-c\sqrt{\log x}}\right).$$

Letting $[\sqrt{x}] = K$, and following equation 2.2 the above is then

$$\begin{aligned} \sum_{K < p \leq x} p \left[\frac{x}{p}\right] &= \sum_{n < K-1} n \left(\sum_{\frac{x}{n+1} < p \leq \frac{x}{n}} p \right) = \mathcal{K}(x) + \mathcal{K}\left(\frac{x}{2}\right) + \cdots + \mathcal{K}\left(\frac{x}{K-1}\right) - K\mathcal{K}\left(\frac{x}{K}\right) \\ &= \sum_{n < K-1} \left(\mathcal{K}\left(\frac{x}{n}\right) - \mathcal{K}\left(\frac{x}{K}\right) \right). \end{aligned}$$

and hence by 4.1

$$\begin{aligned}
 \sum_{n \leq x} P(n) &= \sum_{n < \sqrt{x}} \int_{\sqrt{x}}^{\frac{x}{n}} \frac{t}{\log t} dt + \sum_{n < \sqrt{x}} O\left(\frac{x^2}{n^2} e^{-c\sqrt{\log \frac{x}{n}}}\right) \\
 (4.2) \quad &= \int_{\sqrt{x}}^x \frac{t \left[\frac{x}{t}\right]}{\log t} dt + O\left(x^2 e^{-c\sqrt{\log x}}\right)
 \end{aligned}$$

$$(4.3) \quad = x \operatorname{li}_g(x) + O\left(x^2 e^{-c\sqrt{\log x}}\right),$$

where the last equality follows since $\int_2^{\sqrt{x}} \frac{t \left[\frac{x}{t}\right]}{\log t} dt = O\left(x^{\frac{3}{2}}\right)$. Turning our attention to this integral function, we make the substitution $t = \frac{x}{u}$ to get

$$(4.4) \quad \int_{\sqrt{x}}^x \frac{t \left[\frac{x}{t}\right]}{\log t} dt = x^2 \int_1^{\sqrt{x}} \frac{[u]}{u^3 \log\left(\frac{x}{u}\right)} du,$$

Setting $d_k = \frac{1}{k!} \int_1^\infty \frac{[u]}{u^3} (\log u)^k du$, and by following a proof identical to that of Lemma 6, we have that

$$(4.5) \quad \int_{\sqrt{x}}^x \frac{t \left[\frac{x}{t}\right]}{\log t} dt = d_0 \frac{x^2}{\log x} + d_1 \frac{1! x^2}{(\log x)^2} + \cdots + d_{k-1} \frac{(k-1)! x^2}{(\log x)^k} + O\left(\frac{1}{(\log x)^{k+1}}\right).$$

Similar to equation 2.5, integration by parts tells us that for $s > 1$

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \int_1^{\infty} x^{-s} d[x] = s \int_1^{\infty} [x] x^{-s-1} dx$$

so that exponential generating series yield

$$\sum_{k=0}^{\infty} d_k z^k = \int_1^{\infty} [x] x^{z-3} dx = \frac{\zeta(2-z)}{2-z}.$$

Using the Taylor series for $\zeta(2-z)$, this is

$$\left(\sum_{j=0}^{\infty} (-1)^j \zeta^{(j)}(2) \frac{z^j}{j!} \right) \left(\frac{1}{2} \sum_{k=0}^{\infty} \frac{z^k}{2^k} \right) = \frac{1}{2} \sum_{n=0}^{\infty} z^n \sum_{k+j=n} (-1)^j \frac{\zeta^{(j)}(2)}{j!} \frac{1}{2^k}$$

so we see that

$$d_n = \frac{1}{2^{n+1}} \sum_{j=0}^n \frac{2^j (-1)^j \zeta^{(j)}(2)}{j!},$$

which establishes Theorem 3. Lastly, we note that

$$\frac{(-1)^j \zeta^{(j)}(2)}{j!} = \frac{1}{j!} \sum_{k=1}^{\infty} \frac{(\log k)^j}{k^2} \approx \frac{1}{j!} \int_1^{\infty} \frac{(\log x)^j}{x^2} dx.$$

A substitution shows that this last integral is the Gamma function, and that it evaluates to exactly $j!$. This allows us to prove that

$$\frac{(-1)^j \zeta^{(j)}(2)}{j!} - 1 \rightarrow 0$$

as $j \rightarrow \infty$, and from this we deduce that as $n \rightarrow \infty$

$$d_n \rightarrow \frac{1}{2^{n+1}} \sum_{j=0}^n 2^j \rightarrow 1.$$

Acknowledgments

I would like to thank Qiaochu Yuan for suggesting the use of generating series, and providing the proof of Lemma 7 on Math Overflow. I would also like to thank Andrew Granville for his many insightful comments during our conversations. I am most grateful to Greg Martin's for both his mathematical suggestions and encouragement, as this project would not have started or progressed otherwise.

REFERENCES

- [1] K. Alladi and P. Erdős. On an additive arithmetic function. *Pacific J. Math.*, 71(2):275–294, 1977.
- [2] N. G. de Bruijn. On the number of positive integers $\leq x$ and free of prime factors $> y$. *Nederl. Acad. Wetensch. Proc. Ser. A.*, 54:50–60, 1951.
- [3] Jean-Marie De Koninck and Aleksandar Ivić. The distribution of the average prime divisor of an integer. *Arch. Math. (Basel)*, 43(1):37–43, 1984.
- [4] Persi Diaconis. Asymptotic expansions for the mean and variance of the number of prime factors of a number n . 1976.
- [5] Aleksandar Ivić. On the k th prime factor of an integer. *Zb. Rad. Prirod.-Mat. Fak. Ser. Mat.*, 20(1):63–73, 1990.
- [6] Éric Saias. Sur le nombre des entiers sans grand facteur premier. *J. Number Theory*, 32(1):78–99, 1989.
- [7] Paul Turán. On a Theorem of Hardy and Ramanujan. *J. London Math. Soc.*, S1-9(4):274.

E-mail address: naslund.eric@gmail.com